

September 2026

Gesche Brandt | Kolja Briedis | Susanne de Vogel | Steffen Jaksztat | Anne-Marie Lapstich | Tina Oestreich | Henrike Steins | Carola Teichmann | Sandra Vietgen | Marten Wallis

DZHW-Promoviertenpanel 2014

Daten- und Methodenbericht zu den Erhebungen der Promoviertenkohorte 2014 (Befragungswelle 1-10);
Version 6.0.0

Dieses Werk steht unter der Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Autor*innen

Gesche Brandt
Kolja Briedis
Susanne de Vogel
Steffen Jaksztat
Anne-Marie Lapstich
Tina Oestreich
Henrike Steins
Carola Teichmann
Sandra Vietgen
Marten Wallis

Impressum

Herausgegeben von

Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)
Lange Laube 12 | 30159 Hannover | www.dzhw.eu
Postfach 2920 | 30029 Hannover
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

Geschäftsführung

Dr. Marcus Beiner
Regina Oelfke

Vorsitzender des Aufsichtsrats

Ministerialdirigent Peter Greisler

Registergericht

Amtsgericht Hannover | B 210251
Umsatzsteuer-Identifikationsnummer:
DE291239300

September 2026

Inhaltsverzeichnisüberschrift

Tabellen-/Abbildungsverzeichnis	II
1 Inhalt und Anlage der Studie	3
2 Erhebungsinstrumente	4
2.1 Pretest.....	4
2.2 Kerninstrumente	4
2.2.1 Erwerbstätigkeit	4
2.2.2 Berufliche und wissenschaftliche Weiterqualifikation	5
2.2.3 Wissenschaftliche Aktivitäten	5
2.2.4 Internationale Mobilität	5
2.2.5 Private Lebenssituation	5
2.3 Lernumwelt und Rahmenbedingungen	6
2.4 Wellenspezifische (Schwerpunkt-)Themen.....	6
2.4.1 Internationale Mobilität	6
2.4.2 Individuelles Innovationsverhalten	6
2.4.3 Persönlichkeitsmerkmale	6
2.4.4 Gesundheit und Beeinträchtigung	6
2.4.5 Politische Partizipation	7
2.4.6 Fragen zu COVID-19.....	7
3 Grundgesamtheit und Stichprobenverfahren	8
3.1 Grundgesamtheit	8
3.2 Vollerhebung.....	8
4 Durchführung der Erhebung	9
4.1 Erstbefragung.....	9
4.2 Folgebefragungen	9
5 Rücklauf	11
6 Datenaufbereitung	14
6.1 Datenübertragung.....	14
6.2 Codierung offener Angaben.....	14
6.3 Datenprüfung und Datenbereinigung	16
6.4 Generierung von Variablen	17
6.5 Erstellung der Datensätze	17
6.6 Vergabe von Variablennamen, Variablen- und Wertelabels	18

6.7	Codierung fehlender Werte	20
7	Gewichtung	22
7.1	Vorgehen und Anwendungshinweise	22
7.2	Modellierung der Ausfallgewichte	23
7.3	Kalibrierung	24
7.4	Trimmung	25
7.5	Verteilung der Gewichte	25
8	Anonymisierung	26
9	Anhang	28
9.1	Logit-Regressionen zur Erstellung der Panelausfallgewicht	28
9.2	Dokumentation der Herkunft sekundär genutzter Quellen	39
10	Literaturverzeichnis	40

Tabellen-/Abbildungsverzeichnis

Tabelle 1:	Brutto- und Nettorücklaufquote des DZHW-Promoviertenpanels 2014	11
Tabelle 2:	Beteiligung der Hochschulen	12
Abbildung 1:	Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf Welle 2 bis Welle 7 13	
Tabelle 3:	Vercodete Merkmale und verwendete Codierlisten	15
Tabelle 4:	Systematik der Themenbereiche und Variablenstämme	18
Tabelle 5:	Systematik des FDZ-DZHW für fehlende Werte	20
Tabelle 6:	Bereitgestellte Gewichte zum DZHW-Promoviertenpanel 2014	23
Tabelle 7:	Verteilung der Gewichte für Welle 1 bis 7	25
Tabelle 8:	Zugangswege im FDZ-DZHW	26
Tabelle 9:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 3	28
Tabelle 10:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 4	30
Tabelle 11:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 5	32
Tabelle 12:	Logit-Regression zur Erstellung des Ausfallgewichts für Welle 6	34

1 Inhalt und Anlage der Studie

Das DZHW-Promoviertenpanel startete im Jahr 2013 im Rahmen der Förderlinie „Forschung zum wissenschaftlichen Nachwuchs“ (FoWiN) des Bundesministeriums für Bildung und Forschung¹. Im Rahmen der Studie wird untersucht, welche Einflüsse die formalen Promotionskontexte und die konkreten Lern- und Entwicklungsbedingungen, die Promovierte während ihrer Promotionsphase vorgefunden haben, auf den beruflichen Übergang nach der Promotion und auf den weiteren Berufsverlauf innerhalb und außerhalb der Wissenschaft ausüben. Die Grundgesamtheit der Befragung umfasst alle 28.147 Personen (Statistisches Bundesamt 2015), die im Prüfungsjahr 2014 in Deutschland an einer Hochschule mit Promotionsrecht eine Promotion abgeschlossen haben. Die Befragung ist als Vollerhebung angelegt, um eine ausreichend große Fallzahl zu erhalten (für Details zu den ersten beiden Wellen siehe: Brandt et al. 2017).

Teile der Datenaufbereitung sowie die Übersetzung und Evaluation des Übersetzungsprozesses wurden im Rahmen des vom Bundesministerium für Forschung, Technologie und Raumfahrt finanzierten Projektes „Übersetzungsprozessoptimierung im Kontext der Internationalisierung des FDZ-DZHW am Beispiel zusätzlicher Datenveröffentlichungen (OptiDat)“ durchgeführt.

Die Auftaktbefragung fand im Jahr 2015 statt; hierzu wurde ein Paper-Pencil-Fragebogen circa ein halbes bis anderthalb Jahre nach Abschluss der Promotion postalisch versandt. Die Zweitbefragung des Prüfungsjahrgangs erfolgte etwa ein Jahr später als Onlinebefragung. Die weiteren Folgebefragungswellen erfolgten in jährlichem Abstand jeweils im Frühjahr (Welle 2-10).

Neben Grunddaten für die Bildungsberichterstattung enthält der Datensatz detaillierte Informationen über die Lern- und Entwicklungsbedingungen während der Promotionsphase sowie über beruflichen Werdegänge Promovierter im Anschluss an die Promotion. Darüber hinaus beinhaltet der Datensatz eine Reihe von Persönlichkeitsmerkmalen (Big Five, Selbstwirksamkeit, internale/externale Kontrollüberzeugungen) sowie sozio- und bildungsbiografische Hintergrundinformationen. Dadurch bietet sich ein großes, bisher nicht vorhandenes Analysepotential für die Hochschul- und Wissenschaftsforschung. Das Paneldesign und die Erhebung von monatsgenauen Verlaufsdaten ermöglichen Kausalanalysen auf der Individualebene im Zeitverlauf (z. B. in Form von Ereignisdaten- und Sequenzmusteranalysen).

¹ Förderkennzeichen 16FWN014 und 16FWN017

2 Erhebungsinstrumente

In der ersten Befragungswelle des Promoviertenpanels 2014 wurde als Erhebungsinstrument ein standardisierter Papierfragebogen in deutscher Sprache eingesetzt.² Für die Befragungswellen zwei bis zehn wurden zur Datenerhebung standardisierte Online-Fragebögen in deutscher Sprache verwendet.³ Zu den Kerninstrumenten, die in jeder Welle erneut abgefragt werden, gehören Fragen zu den Themen Erwerbstätigkeit, berufliche und wissenschaftliche Weiterqualifikationen, wissenschaftliche Aktivitäten, Mobilitätserfahrungen und Informationen zur privaten Lebenssituation. Die einzelnen Erhebungswellen enthalten darüber hinaus individuelle Schwerpunktsetzungen. Die Programmierung des Fragebogens und die Durchführung der Befragung erfolgten mit der DZHW-Onlinebefragungssoftware „Zofar“.

Für die Entwicklung der Erhebungsinstrumente wurden einzelne Fragen aus bestehenden Befragungsinstrumenten übernommen beziehungsweise adaptiert. Eine Übersicht der verwendeten Quellen findet sich in Anhang **Fehler! Verweisquelle konnte nicht gefunden werden..**

2.1 Pretest

Vor der Erstbefragung wurden zunächst ein kognitiver und später ein quantitativer Pretest durchgeführt. In diesen Phasen sollten vor allem die neu entwickelten Items zur Erhebung der Lernumwelt in der Promotionsphase getestet und die Anwendbarkeit für die Befragung von Promovierenden und Promovierten verschiedener Fächer und Promotionsformen geprüft werden. Das Ziel der Pretests war es, das Frageverständnis und das Antwortverhalten der Promovierenden und Promovierten zu eruieren, die Beantwortungsdauer zu ermitteln und etwaige Reihenfolgeeffekte aufzudecken. Ausführlich ist dies in einem Werkstattbericht dokumentiert (Brandt, Vogel & Jaksztat, 2016).

Auch der Fragebogen der Zweitbefragung wurde vorab im Rahmen eines kognitiven Pretests mit Promovierten unterschiedlicher formaler Promotionsformen und Fachbereiche getestet.

2.2 Kerninstrumente

Im Nachfolgenden werden die Kerninstrumente hinsichtlich ihrer verschiedenen Themen näher beschrieben.

2.2.1 Erwerbstätigkeit

Die Fragebögen enthalten ab der zweiten Erhebungswelle zwei Instrumente zur Erfassung der Tätigkeitsverläufe im Anschluss an die Promotion bzw. seit der letzten Befragung: Zum einen kam ein Berufstableau zur monatsgenauen Erfassung wichtiger beruflicher Episoden und zur Abbildung detaillierter Merkmale der jeweiligen Episode (Beginn und Ende, berufliche Stellung, Vertragsart, Arbeitszeit, Arbeitsort, wissenschaftliche Tätigkeit und Arbeitgeber) zum Einsatz; darüber hinaus wurde ein Kalendarium eingesetzt, um die beruflichen und nicht-beruflichen Tätigkeiten nach der Promotion monatsgenau zu erfassen (z. B. weitere Ausbildungen, Stipendien, Elternzeit, Familienarbeit, Arbeitslosigkeit). Um die Fortschreibung der Daten über die Wellen zu erleichtern, wurden jeweils alle Tätigkeiten seit der letzten Befragung, an der die Promovierten teilnahmen, erfasst. Befragte, die z. B. nicht an der zweiten Befragung teilgenommen haben, wurden in der dritten Welle um Angaben seit

² Auf Wunsch konnte von den Befragten ein deutscher oder englischer PDF-Fragebogen per E-Mail angefordert werden.

³ Die Fragebögen, Filterführungsdiagramme und Variablenpläne können im Metadatensystem heruntergeladen werden.

Abschluss ihrer Promotion gebeten, um möglichst vollständige Verlaufsdaten zu erhalten. Technisch wurde dies durch die Verwendung von Preloads und die Einblendung alternativer Fragetexte gelöst. Abschließend wurden die Befragten um die Kennzeichnung der Episode gebeten, bei der es sich um ihre aktuelle bzw. letzte Beschäftigung handelt.

Für die aktuelle bzw. letzte Stelle wurde zudem in jeder Welle eine Reihe weiterer Beschäftigungsmerkmale erhoben. Dazu gehören die Berufsbezeichnung, der Tätigkeitsbereich und die beruflichen Aufgaben, anhand derer während der Datenaufbereitung eine Vercodung nach der Klassifikation der Berufe 2010 (KldB2010) für die Wellen 1 bis 5 vorgenommen wurde⁴. Des Weiteren wurde erfragt, ob es sich um eine Tätigkeit im öffentlichen Dienst handelt, und die Betriebsgröße, die Branche, die berufliche Zufriedenheit, die Qualifikationsadäquanz der Beschäftigung und das Einkommen erhoben. Ab der dritten Welle wurde bei Tätigkeiten in der Wissenschaft erstmalig auch die Personalkategorie erhoben. Schließlich wurden in jeder Welle Fragen zum Thema Selbständigkeit gestellt.

2.2.2 Berufliche und wissenschaftliche Weiterqualifikation

In der zweiten und dritten Welle wurden Informationen zu berufsständisch organisierten Weiterbildungen gesammelt. Da sich derartige Weiterbildungen üblicherweise über mehrere Jahre erstrecken, wurden Details zur berufsständischen Weiterbildung danach noch in der fünften Befragung erhoben. Zugleich wurde der Themenblock ab der dritten Befragung um zwei Fragen zum Abschluss der Weiterqualifikation ergänzt, die jährlich abgefragt werden.

Zum Kerninstrumentarium für die Erhebung ggf. stattfindender wissenschaftlicher Weiterqualifikation gehören Fragen zur Art und zum Fortschritt der Weiterqualifizierung, zur wissenschaftlichen Karriereintention, zur Einschätzung der Karriereaussichten und detailliertere Informationen zur Intention bzw. Art der Professur.

2.2.3 Wissenschaftliche Aktivitäten

Ab der zweiten Befragung wurden Informationen zur Wissenschaftshaltigkeit der Arbeit erfragt. Um die Befragungslast für die Befragten so gering wie möglich zu halten, erfolgte die detaillierte Abfrage wissenschaftlicher Aktivitäten ab der dritten Befragung nach einer einleitenden Filterfrage. Jährlich erfragt wurden Publikationen, Tagungsbesuche, Gutachter*innentätigkeiten, Drittmittelwerbung, Patente, Lehrtätigkeiten und die subjektive Integration in die Scientific Community. In der fünften Welle wurden zudem zwei Instrumente zur Erfassung der Mentoringsituation in der Postdoc-Phase eingesetzt. Die Mitarbeit in akademischer Selbstverwaltung sowie die Mitgliedschaft in Verbänden und Fachgesellschaften wurden ab der zweiten Welle nicht mehr erfragt.

2.2.4 Internationale Mobilität

Seit der dritten Welle gehören Fragen zu internationalen Mobilitätserfahrungen und -plänen zum jährlichen Frageprogramm. Erfragt werden die absolvierten Auslandsaufenthalte seit der letzten Befragung und die Pläne, zukünftig ins Ausland zu gehen.

2.2.5 Private Lebenssituation

In diesem Themenbereich wurden in jeder Welle Informationen zur Partnerschaftssituation und Elternschaft erhoben. Dazu gehören beispielsweise der Familienstand, ggf. Angaben zum Bildungsstand des Partners/der Partnerin, zur Erwerbstätigkeit und zum Tätigkeitsbereich der Partner*innen sowie zur Haushaltssituation. Die Fragen nach vorhandenen Kindern unterscheiden sich in der dritten und vierten Welle im Vergleich zu den Vorwellen dahingehend, dass nur noch nach Kindern gefragt wurde, die seit der letzten Teilnahme dazu gekommen sind. Ab der fünften Welle wurden wieder die Gesamtzahl der Kinder und alle Geburtsdaten erhoben. Ebenso wurden in der zehnten Welle

⁴ Ab Welle 6 wurde aus Kostengründen auf die aufwändige Kodierung der Berufe verzichtet.

Pflegeverantwortungen und detaillierte Informationen zu eventuellen Unterbrechungen der Erwerbstätigkeit aufgrund von Kindern erfragt.

2.3 Lernumwelt und Rahmenbedingungen

Zur Erfassung der Lernumwelt in der Promotionsphase wurde für die erste Welle ein theoretisch fundiertes und empirisch getestetes Modell entwickelt. Die theoretische Konzeption der Lernumwelten basiert auf dem SSCO-Modell (*Structure - Support - Challenge - Orientation*) von Bäumer, Preis, Roßbach, Stecher & Klieme (2011). Details zum Instrument sind in den dazu entstandenen Publikationen nachzulesen (Brandt et al. 2016; de Vogel et al. 2017).

2.4 Wellenspezifische (Schwerpunkt-)Themen

In der dritten Welle wurden bei Personen, die nicht an der zweiten Welle teilgenommen haben, zentrale Fragen aus der zweiten Welle nacherhoben. Dazu gehören z. B. Informationen zur Promotion (Veröffentlichung, Datum der Urkundenverleihung, erhaltene Preise), zur internationalen Mobilität vor und während des Studiums und zur Staatsangehörigkeit. Darüber hinaus wurden in jeder Welle spezifische (Schwerpunkt-)Themen gesetzt. In der siebten Welle wurden außerdem Informationen zur Promotionsbetreuung und den Hauptgutachter*innen nacherhoben (jeweils Geschlecht, Position, Sympathie).

2.4.1 Internationale Mobilität

In der dritten Welle wurde ein Schwerpunkt auf internationale Mobilitätserfahrungen gelegt. Neben der Nacherhebung der Auslandsaufenthalte vor dem Promotionsabschluss und neben den Kerninstrumenten zur internationalen Mobilität wurden auch die zukünftigen Mobilitätspläne sowie die Nutzung von Informations- und Kommunikationstechnologien (IKT) für den internationalen Austausch erhoben. Die Nutzung von IKT wurde in der siebten und zehnten Welle ebenfalls erfragt.

2.4.2 Individuelles Innovationsverhalten

Der Fragebogen der vierten Welle enthielt mehrere etablierte Instrumente zum Thema individuelles Innovationsverhalten: Erhoben wurden die Dynamik des beruflichen Umfeldes (deutschsprachige Adaption von Schilke (2014)), Problemlösung und Entscheidungsautonomie im Beruf (Morgeson und Humphrey (2006), deutschsprachige Adaption von Stegmann et al. (2010)) sowie die individuelle Ambidexterität (deutsche Adaption von Kobarg et al. (2017)).

2.4.3 Persönlichkeitsmerkmale

Die Instrumente zur Erfassung zentraler Persönlichkeitsmerkmale, die bereits in die Erstbefragung integriert wurden, wurden in der vierten Welle erneut eingesetzt. Dazu zählen die psychometrischen Kurzskalen zur Messung der allgemeinen Selbstwirksamkeit (ASKU) (Beierlein et al. 2014b), zur Messung der „Big Five-Persönlichkeitsfaktoren“ (BFI-10) (Rammstedt et al. 2014) sowie internaler und externaler Kontrollüberzeugungen (IE-4) (Kovaleva et al. 2014). ASKU und BFI-10 wurden in der siebten und zehnten Welle ebenfalls abgefragt.

2.4.4 Gesundheit und Beeinträchtigung

Thematischer Schwerpunkt der fünften Welle war der Bereich Gesundheit und Behinderung. Hierzu kam ein Kurzinstrument zur Erfassung des subjektiven allgemeinen Gesundheitszustands zum Einsatz. Gestellt wurden außerdem drei Fragen zur Erhebung der körperlichen und psychischen Gesundheit sowie gesundheitsbedingten Beeinträchtigungen im Alltag. Bei diesen Fragen handelt es sich um eine deutschsprachige Adaption des Healthy Days Core Module (CDC HRQOL– 4) (Centers for Disease Control and Prevention (CDC) 2000), die beispielsweise auch in der Startkohorte 5 des NEPS (Nationales Bildungspanel (NEPS) 2018) eingesetzt wurden. Schließlich wurden Informationen zu Behinderungen

und ggf. die Art, das Jahr der Feststellung und der Grad der Behinderung abgefragt. Diese Fragen sind Instrumente, die bereits in der Startkohorte 6 des NEPS (Nationales Bildungspanel (NEPS) 2013) genutzt wurden. Alle gesundheitsbezogenen Instrumente sind auch Teil des Nacaps-Frageprogramms (National Academics Panel Study). Da es sich bei Daten zum Gesundheitszustand nach DSGVO um besonders schützenswerte Daten handelt, wurde für diese Fragen noch einmal das explizite Einverständnis der Befragten eingeholt.

2.4.5 Politische Partizipation

In der sechsten Welle wurden das politische Interesse und das politische Engagement der Promovierten erhoben. Die sechs Fragen orientierten sich eng an Fragen, die auch in Befragungen des Projekts Eurograduate verwendet wurden und aus dem European Social Survey (ESS) 2014 (Frage 5.2-5.5) bzw. ESS 2016 (Frage 5.6) stammen. Frage 5.1, in der das Politikinteresse auf einer fünfstufigen Skala erfragt wurde, wurde mit einem Methodenexperiment verbunden. Hintergrund war, dass die Skalenrichtung im Originalinstrument nicht der üblichen Skalenrichtung der Promoviertenbefragung entsprach. Daher wurde allen Befragten nach dem Zufallsprinzip eine der beiden Skalenrichtungen angezeigt (Version 1: von 1 „sehr interessiert“ bis 5 „überhaupt nicht interessiert“, Version 2: von 1 „überhaupt nicht interessiert“ bis 5 „sehr interessiert“), um zu prüfen, ob sich die Antworten unterscheiden. Der Datensatz enthält eine generierte Variable, in der alle Befragten die gleiche Skalenrichtung aufweisen, indem die Antworten einer der beiden Gruppen umgepolt wurde (k6polinter).

2.4.6 Fragen zu COVID-19

In der siebten und achten Erhebungswelle wurden Fragen zur Beeinträchtigung durch die Auswirkungen von COVID-19 erfragt. Die Fragen stammen aus einer Zusatzbefragung des NEPS (Startkohorte 2-6) im Mai 2020. Es wurde erfragt, ob sich bedingt durch die COVID-19-Pandemie die Arbeitssituation verändert hat, ob sich die Auftrags- bzw. Einkommenssituation von Selbständigen geändert hat und wie stark verschiedene Lebensbereiche betroffen waren.

3 Grundgesamtheit und Stichprobenverfahren

3.1 Grundgesamtheit

Als wissenschaftlicher Nachwuchs werden in Anlehnung an die Definition in der Berichterstattung des Bundes Hochschulabsolvent*innen verstanden, die nach dem Studium innerhalb oder außerhalb von Hochschulen oder außeruniversitären Forschungseinrichtungen wissenschaftlich tätig sind oder promovieren, sowie Promovierte, die innerhalb und außerhalb von Hochschulen und außeruniversitären Forschungseinrichtungen einer wissenschaftlichen Tätigkeit nachgehen oder nachgehen wollen. Die Grundgesamtheit umfasst alle Promovierten, die im Prüfungsjahr 2014 (Wintersemester 2013/14 und Sommersemester 2014) an einer deutschen Hochschule mit Promotionsrecht eine Promotion abgeschlossen haben. Die amtliche Statistik weist für das betreffende Prüfungsjahr 2014 bundesweit 28.147 Promovierte aus (Statistisches Bundesamt, 2015).

3.2 Vollerhebung

Im ersten Schritt wurden die Hochschulleitungen aller (damals) 146 Hochschulen mit Promotionsrecht über die geplante Erhebung informiert und um Unterstützung des Forschungsvorhabens gebeten. Die Erstansprache der Zielpersonen erfolgte aus Datenschutzgründen über die Hochschulen. Für die Folgebefragungen wurden die Kontaktdaten der Teilnehmenden erfasst und zwischen den Erhebungswellen (7 Erhebungswellen) fortlaufend aktualisiert. Die Ausgangsstichprobe der Panelbefragung umfasste alle Befragten, die einer weiteren Teilnahme zugestimmt hatten und für die gültige Kontaktdaten vorlagen. Die Zahl der kontaktierten Personen verringerte sich in den Folgewellen aufgrund von Widerrufen und ungültig gewordenen Kontaktdaten.

4 Durchführung der Erhebung

4.1 Erstbefragung

[Kontaktaufnahme und Adresspflege] Die Erstansprache der Promovierten erfolgte aus Datenschutzgründen über die Dekanate bzw. die für Promotionen zuständigen Prüfungsämter. In den meisten Fällen lag die Verantwortlichkeit innerhalb der einzelnen Dekanate. An wenigen Hochschulen wurden zentrale bzw. dezentrale Prüfungsämter oder Graduiertenakademien als zuständige Stellen identifiziert.

[Erhebungsunterlagen] Die Erhebungsunterlagen bestanden in der ersten Befragungswelle pro zu befragende Person aus einem Anschreiben (inkl. Datenschutzinformationen), dem Fragebogen (Papier-Pencil), einem Begleitschreiben des Bundesministeriums für Bildung und Forschung (BMBF) und einem an das DZHW adressierten portofreien Umschlag zur Rücksendung des ausgefüllten Fragebogens. Zur Erhöhung der Teilnahmebereitschaft wurden zwei Erinnerungsschreiben ebenfalls postalisch versandt.

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich vom 15. Dezember 2014 bis zum 17. Februar 2016. Die beiden Erinnerungsschreiben wurden etwa vier bzw. acht Wochen nach Feldstart verschickt. Dieser lange Zeitraum ist dadurch begründet, dass die Verwaltungen der Hochschulen unterschiedlich schnell auf die Anfrage des Projektes reagiert haben. Nicht alle Verwaltungsstellen konnten für den Versand der Erinnerungen gewonnen werden. Bei dem Großteil der Verwaltungsstellen konnte der Versand der Befragungsunterlagen und der zwei Erinnerungsschreiben jedoch planmäßig durchgeführt werden. Aufgrund des angewendeten Kontaktverfahrens über die Prüfungsämter konnte das DZHW keinen direkten Einfluss auf den genauen Versandzeitpunkt der Erhebungsunterlagen nehmen.

[Rücklaufsteigernde Maßnahmen] Die rücklaufsteigernden Maßnahmen zielten zum einen auf Anreize für die Hochschulen, die Befragung organisatorisch zu unterstützen und zum anderen auf die individuelle Motivation der Befragten. Als Dank für die Unterstützung erhielten die Hochschulen exklusive Vorabergebnisse aus der Befragung. Die Fakultäten konnten auf Wunsch Sonderauswertungen für die eigene Hochschule erhalten, sofern dafür ausreichende Fallzahlen vorlagen. Die auf die individuelle Motivation der Befragten zielenden rücklaufsteigernden Maßnahmen bestanden darin, dass zunächst in dem Anschreiben an die Befragten die subjektive und gesellschaftliche Relevanz des Themas hervorgehoben wurde. Zusätzlich stellte das Begleitschreiben des Bundesministeriums für Bildung und Forschung (BMBF) in der ersten Welle die hochschulpolitische Relevanz der Studie dar und warb um Unterstützung. Des Weiteren erfolgte der Versand von Erinnerungsschreiben. Mittels einer Rückmeldung ausgewählter Studienergebnisse nach der ersten Welle sollte zum einen die Bindung der Studienteilnehmenden an das Panel erhöht werden, um diese zur Teilnahme an Folgewellen zu motivieren. Zum anderen diente der Ergebnisversand zur erneuten Aktualisierung des Adressbestandes. Darüber hinaus wurden unter allen Teilnehmenden der ersten Welle ein Laptop, drei iPads sowie mehrere Büchergutscheine verlost.

4.2 Folgebefragungen

[Kontaktaufnahme und Adresspflege] Für die Folgebefragungen wurde auf Grundlage der in der Erstbefragung erhobenen Kontaktdaten ein Adressdatensatz aufgebaut. Zur Wahrung der Datenschutzbestimmungen wurden Befragungs- und Adressdaten über eine Identifikationsnummer verknüpft und

getrennt voneinander gespeichert. Die Adressbestände wurden zwischen den Erhebungswellen kontinuierlich aktualisiert. Zusätzlich wurden Projektergebnisse, Ankündigungen kommender Befragungen sowie ab einzelnen Wellen auch postalische Kontaktierungen (z.B. Vorankündigungsschreiben oder Weihnachtspostkarten) genutzt, um die Panelbindung zu stärken und die Aktualität der Kontaktdaten zu überprüfen. Ab der sechsten Welle erfolgte die Kontaktaufnahme ausschließlich per E-Mail.

[Erhebungsunterlagen] Die Folgebefragungen wurden als Online-Befragungen mit der DZHW-Onlinebefragungssoftware „Zofar“ durchgeführt. Die Einladungen erfolgten überwiegend per E-Mail und enthielten einen personalisierten Zugangslink zur Befragung, Informationen zum Datenschutz sowie Hinweise auf die angekündigten Incentives. In einzelnen Wellen wurden ergänzend postalische Vorankündigungen oder Zwischeneinladungen versandt. Nicht teilnehmende Personen erhielten jeweils bis zu drei Erinnerungsschreiben. Mit Einführung der Datenschutz-Grundverordnung wurde die Zustimmung zu den Datenschutzbestimmungen über ein Opt-in-Verfahren eingeholt.

[Feldphase] Die Folgebefragungen wurden jährlich im Zeitraum von 2016 bis 2024 durchgeführt. Die Feldphasen erstreckten sich jeweils über mehrere Wochen. Nach Versand der Einladungsschreiben wurden gestaffelt bis zu drei Erinnerungsschreiben an Personen versandt, die noch nicht an der jeweiligen Erhebung teilgenommen hatten. Die genauen Daten der jeweiligen Feldphasen können dem Metadaten-System des FDZ-DZHW entnommen werden.

[Rücklaufsteigernde Maßnahmen] Zur Förderung der Teilnahme wurden verschiedene Maßnahmen zur Panelpflege und Rücklaufsteigerung eingesetzt. Hierzu gehörten regelmäßige Informationen über Projektergebnisse, die Bereitstellung einer Projektwebsite, Vorankündigungen der Befragungen sowie postalische Kontaktaktionen. Darüber hinaus wurden in allen Folgebefragungen Sachpreise und Gutscheine verlost. Die Erinnerungsschreiben sowie die kontinuierliche Kommunikation mit den Teilnehmenden dienten zugleich der Stärkung der Panelbindung und der Sicherung langfristiger Teilnahmebereitschaft.

5 Rücklauf

Tabelle 1 gibt einen Überblick über die Anzahl der versandten Befragungsunterlagen bzw. Einladungen sowie die Rücklaufquoten der einzelnen Erhebungswellen. In der Erstbefragung wurden rund 19.900 Fragebögen über die beteiligten Hochschulen versandt. Insgesamt gingen 5.423 Fragebögen ein⁵, von denen nach Bereinigung um leere, doppelte und unlesbare Fragebögen sowie nachträgliche Panelbereinigungen 5.408 Fälle in den Befragungsdatensatz aufgenommen wurden.

Für die zweite Befragungswelle erklärten sich 4.816 Personen (89% der Teilnehmenden der Erstbefragung) mit einer erneuten Kontaktaufnahme einverstanden. Die Rücklaufquoten der Folgebefragungen lagen zwischen 55,6% und 66,1% und blieben bis zur fünften Welle weitgehend stabil. In den Wellen sechs und sieben gingen sie auf rund 56% zurück und blieb daraufhin weitestgehend stabil. Über alle Erhebungswellen hinweg zeigte sich ein deutlicher Anstieg des Rücklaufs nach Versand der Erinnerungsschreiben, was deren Wirksamkeit zur Steigerung der Teilnahmebereitschaft unterstreicht.

[Panelausfälle] Im Verlauf des Panels kam es zu den für Längsschnittstudien typischen Panelausfällen. Diese resultierten insbesondere aus aktiven Widerrufen der Teilnahme, ungültig gewordenen Post- oder E-Mail-Adressen sowie vereinzelt aus doppelten Teilnahmen⁶. Darüber hinaus traten in allen Folgebefragungen sogenannte „weiche“ Verweigerungen auf, bei denen der Onlinefragebogen abgerufen, jedoch ohne inhaltliche Beantwortung wieder verlassen wurde. Solche Fälle wurden nicht als endgültige Panelausfälle gewertet, da die betreffenden Personen in den folgenden Wellen erneut kontaktiert wurden. Die Zahl der zur Teilnahme eingeladenen Personen verringerte sich im Verlauf der Panelstudie kontinuierlich durch Widerrufe und nicht mehr erreichbare Teilnehmende. Ab der sechsten Erhebungswelle erfolgte die Kontaktaufnahme ausschließlich per E-Mail, wodurch Personen ohne gültige E-Mail-Adresse nicht mehr in die Befragung einbezogen werden konnten. Trotz dieser paneltypischen Ausfälle konnte über den gesamten Erhebungszeitraum ein stabiler Rücklauf erzielt werden.

Tabelle 1: Brutto- und Nettorücklaufquote des DZHW-Promoviertenpanels 2014

	Welle 1	Welle 2	Welle 3	Welle 4	Welle 5	Welle 6	Welle 7
Befragungsunterlagen versendet	19.916	4.816	4.809	4.795	4.778	4.558	4.526
davon Versand bestätigt	19.900 ⁷	4.815	4.806	4.789	4.771	4.524	4.500
Befragungsunterlagen Rücklauf	5.423	3.334	2.976	3.037	3.085	2.585	2.625
Befragungsunterlagen verwertbar	5.408	3.183	2.924	2.981	3.037	2.535	2.574
Rücklaufquote							

⁵ Die verwertbaren Befragungsunterlagen umfassen alle Promovierten, die in die Teilnahme eingewilligt haben und mindestens eine Antwort im Onlinefragebogen gegeben haben.

⁶ Vereinzelt lieferten Befragte über mindestens zwei Wellen zwei Fragebögen, was zu identischen Fällen mit unterschiedlichen IDs führte. Nach der jeweiligen Identifikation wurde vom Primärforschungsprojekt ein Fall rückwirkend gelöscht.

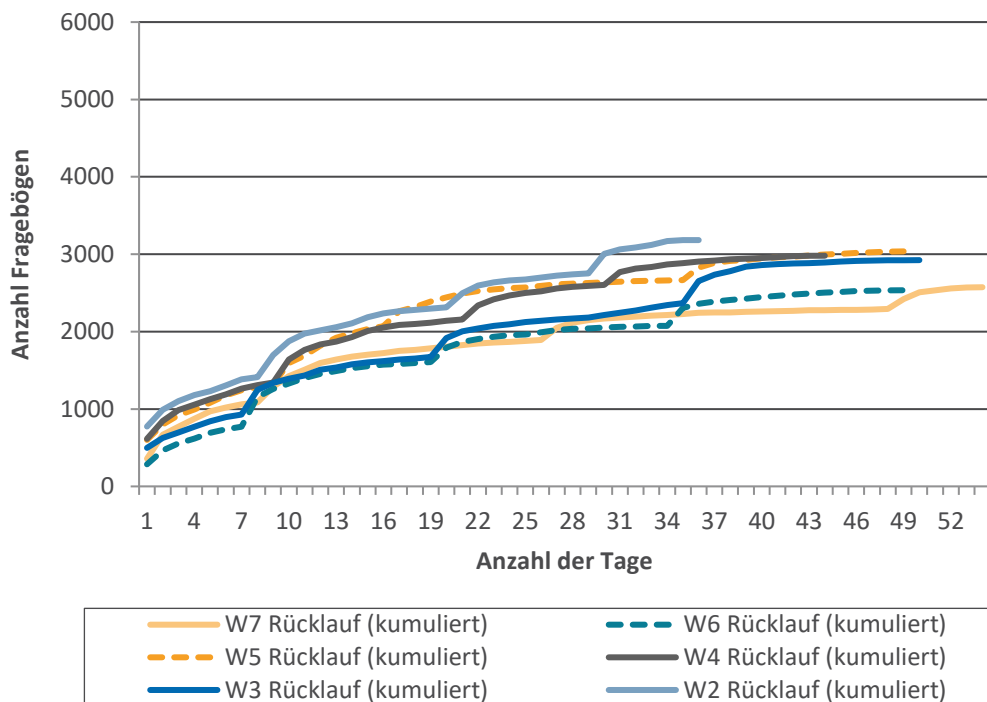
⁷ Aufgrund fehlender Informationen kann die Tabelle für Welle 8-10 zu diesem Zeitpunkt nicht ergänzt werden.

(verwertbar/versendet)	27,2 %	66,1 %	60,8 %	62,2 %	63,6 %	55,6 %	56,9 %
Anteil an Welle 1 (verwertbare Fälle/verwertbare Fälle Welle 1)	-	58,9 %	54,1 %	55,1 %	56,2 %	46,9 %	47,6 %

Tabelle 2: Beteiligung der Hochschulen

Beteiligte	Anzahl
Hochschulen mit Promotionsrecht	146
Keine abgeschlossenen Promotionen im Prüfungsjahr 2014	19
Teilgenommen	112
➤ vollständig teilgenommen	80
➤ teilweise teilgenommen	32
nicht teilgenommen	15

Abbildung 1: Rücklauf des DZHW-Promoviertenpanels 2014 im Zeitverlauf Welle 2 bis Welle 7



In Abbildung 1 ist der Rücklauf der zweiten bis siebten Befragung des Promoviertenpanels im zeitlichen Verlauf dargestellt. Im Vergleich zur dritten Welle weisen die Kurven der zweiten, vierten und fünften Welle einen steileren Verlauf sowohl kurz nach Befragungsstart als auch nach den einzelnen Erinnerungsaktionen auf. In der vierten und fünften Welle wurde jeweils vor Beginn der Befragung ein postalisches Vorankündigungsschreiben verschickt. Bei der dritten Welle wurde dieses während der Befragung versendet. Die Rückläufe dieser Befragungsrunden nähern sich zum Befragungsende an, jedoch liegen die Rücklaufzahlen der vierten und fünften Welle leicht über denen der dritten Welle.

Die Verläufe der zweiten bis zur fünften Welle ähneln sich stark, doch ab der sechsten Welle fällt der Rücklauf erkennbar niedriger aus.

6 Datenaufbereitung

Die im Folgenden beschriebenen Schritte der Datenaufbereitung erfolgten im Anschluss an die Datenerhebung. Die in den Kapiteln 6.1 bis 6.3 beschriebenen Arbeitsschritte – Datenübertragung, Codierung offener Angaben, Datenprüfung und Datenbereinigung – wurden von dem Primärforschungsprojekt durchgeführt. Für die Wellen 8 bis 10 wurden Teile dieser Arbeiten im Rahmen des vom Bundesministerium für Forschung, Technologie und Raumfahrt finanzierten Projektes „Übersetzungsprozessoptimierung im Kontext der Internationalisierung des FDZ-DZHW am Beispiel zusätzlicher Datenveröffentlichungen (OptiDat)“ durchgeführt.

Die Generierung der Variablen (Kapitel 6.4) wurde sowohl von dem Primärforschungsprojekt als auch im Rahmen der Datenedition des FDZ-DZHW vorgenommen. Die Erstellung der Datensätze, Vergabe von Variablennamen, Variablen- und Wertelabels sowie die Codierung fehlender Werte (Kapitel 6.5 bis 6.7) wurden vom Primärforschungsteam in Absprache mit dem FDZ erstellt. Die Schritte der Gewichtung und Anonymisierung werden in den Kapiteln 7 und 8 gesondert ausgeführt.

6.1 Datenübertragung

In der ersten Welle wurden die Angaben der Befragten aus den Papierfragebögen auf Basis eines Codeplans zur weiteren Verarbeitung in ein computerlesbares Format übertragen. Zuvor wurden auf den Papierfragebögen numerische Codierungen für die meisten offenen Angaben vermerkt (vgl. Kapitel 6.2) sowie manuelle Vorkorrekturen zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.3).

[Erstellung eines Codeplans] Auf Basis des Fragebogens der Erstbefragung wurde ein Codeplan erstellt. Dabei wurde vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist und welchen Namen diese Variable trägt.

[Datenerfassung] Für die Datenübertragung wurden der Codeplan, weitere Anweisungen zur Datenerfassung sowie die vorbereiteten Papierfragebögen an einen externen Dienstleister übergeben. Die Erfassung der Angaben erfolgte dort manuell durch Schreibkräfte.

Ab der zweiten Befragungswelle wurden die Daten mit dem Online-Befragungstool „Zofar“ erhoben und folglich kein Codeplan oder manuelle Datenerfassung verwendet. Der Rohdatensatz wurde auf einem geschützten Server gespeichert.

6.2 Codierung offener Angaben

[Erstbefragung] Vor der Datenübertragung erfolgte eine Codierung der meisten (halb-)offenen Angaben. Dabei wurden diesen anhand einer Codierliste numerische Codierungen zugeordnet. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Klassifikation der Berufe, Schlüsselverzeichnis der Studenten- und Prüfungsstatistik etc.). Für einige Variablen wurden Codierlisten auf Basis der in den Daten des Promoviertenpanel 2014 vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Einzelne offene Fragen wurden nicht vercodet, weil sie vorwiegend als Kontextinformationen zur Codierung anderer offener Angaben

erfasst wurden. Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden unverändert beibehalten.

[Folgebefragungen] Offene Angaben wurden – wenn möglich – anhand bereits vorliegender Codierschemata aus den Vorwellen und anhand von Klassifikationsschlüsseln der amtlichen Statistik (z. B. Klassifikation der Berufe) codiert. Die Vercodung erfolgte analog zum Vorgehen in Welle 1.⁸

Zur Codierung der beruflichen Tätigkeiten, welche für die Wellen 1 bis 5 in Form von Codes entsprechend der Klassifikation der Berufe 2010 (KldB2010) vorliegt, wurde die Firma STR Coding beauftragt (die Dokumentation zur Codierung ist im Metadatenystem abrufbar). Es wurden neben den offen erhobenen Angaben zur Berufsbezeichnung, dem hauptsächlichen Tätigkeitsbereich und den beruflichen Aufgaben auch die angegebene berufliche Stellung und in Einzelfällen weitere Kontextinformationen, wie der Wirtschaftsbereich (Branche) und das Promotionsfach genutzt. Der vergebene Code bildet das Tätigkeitsfeld und die berufliche Stellung ab. So werden bei gleichbleibender Stelle Veränderungen der Tätigkeiten (wissenschaftlich, administrativ oder beratend etc.) und auch Veränderungen wie der Aufstieg in Führungspositionen abgebildet. Eine Änderung des Berufscode ist demnach aber nicht immer gleichbedeutend mit einem Jobwechsel.

In Tabelle 3 sind die codierten Merkmale sowie die verwendeten Codierlisten der jeweiligen Welle dargestellt. Der Datensatz des Primärforschungsteams beinhaltet die codierten numerischen Variablen sowie die offenen Nennungen. Im SUF hingegen werden zum Datenschutz nur die numerischen Variablen ausgegeben (vgl. Kapitel 8). Die Ausprägungen der einzelnen Variablen sind für den SUF Metadatensuchsystem⁹ dokumentiert (vgl. auch Kapitel 6.6).

Tabelle 3: Vercodete Merkmale und verwendete Codierlisten

Variablenamen ¹⁰	Merkmal	Codierliste
d1prohs	Promotionshochschule, Hochschule des Studienabschlusses	Deutsche Hochschule: Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2013/2014 und SoSe 2014) Ausländische Hochschule: DZHW-Codierung auf Basis der Ländercodes
d1profach1-3, b1studber1-4, b1abs1-4	Promotionsfach, Studienfach, Studienabschluss	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2015/2016 und SoSe 2016)
c1eberuf	Beruf	Destatis Klassifikation der Berufe 2010
c1-5jberuf [c1-5jberuf]	Beruf	Destatis Klassifikation der Berufe 2010
c*taet1plz – c*taet10plz	Ort der beruflichen Tätigkeit	Postleitzahlenverzeichnis
c*taet1land – c*taet6land	Land der beruflichen Tätigkeit	Destatis Staats- und Gebietssystematik 2014
m*studkomplla, m*studsemla, m*studprakla, m*studkursla, m*studsonla	Land des Studienaufenthaltes	Destatis Staats- und Gebietssystematik 2014
m*grad1land, m*grad2land, m*grad3land,	Land des Auslandsaufenthaltes	Destatis Staats- und Gebietssystematik 2014

⁸ Dies betrifft die beruflichen Tätigkeitsbereiche und Aufgaben. Die offenen Angaben dienten als zusätzliche Informationen für die Codierung der ebenfalls offen abgefragten Berufsbezeichnung.

⁹ [https://metadata.fdz.dzhw.eu/#!/de/studies/stu-phd2014\\$](https://metadata.fdz.dzhw.eu/#!/de/studies/stu-phd2014$)

¹⁰ Die angegebenen Variablenamen wurden von der Primärforschungsgruppe vergeben.

m*grad4land		
m*plandaofla, m*plandamfla, m*planzeitla, m*planforla, m*planwbila, m*plansonla	Land des geplanten Auslandsaufenthaltes	Destatis Staats- und Gebietssystematik 2014
k*natausl	Nationalität	Destatis Staats- und Gebietssystematik 2014
c*weibfarzt, c*weibzarzt, c*weibpsycho, c*weibtarzt, c*weibfanw	Sonstige berufsständische Weiterbildung	Einordnung in vorgegebene Kategorien der berufsständigen Weiterbildungen, wenn möglich
c*jbranche	Sonstige Wirtschaftsbereiche	Einordnung in vorgegebene Kategorien der Wirtschaftsbereiche, wenn möglich

Das Zeichen „*“ in den Variablennamen steht für die Wellenkennziffer 1-10.

Für die folgenden Variablen wurde auf eine Codierung offener Angaben verzichtet:

- die Angabe des Promotionspreises (nur in Welle 3 erhoben, Frage 4.3),
- die Tätigkeitsbereiche und beruflichen Aufgaben (Welle 2 bis Welle 5 wurden als Berufscodierung in c1eberuf bzw. c1-5jberuf codiert, ab Welle 6 entfallen diese Codierungen),
- die Kategorie für sonstige Angaben bei der Abfrage der Institution, an der eine Professur angesiedelt ist (nur in Welle 3 erhoben, Frage 2.9),
- die unter „Sonstige“ angegebene Art der Professur (in Welle 3 bis Welle 10 erhoben),
- bei unter „Sonstiges“ offen genannten Personalkategorien (in Welle 2 bis Welle 10 erhoben) und
- der unter „sonstiger“ offen abgefragte berufliche Abschluss des Partners oder der Partnerin (in Welle 2 und Welle 3 erhoben).

6.3 Datenprüfung und Datenbereinigung

[Manuelle Vorkorrektur der Erstbefragung] In der ersten Welle wurden bereits vor der Übertragung der Daten eine manuelle Prüfung und gegebenenfalls eine Anpassung von Angaben der Befragten auf den Papierfragebögen durchgeführt.¹¹ Dies sollte zum einen die Erfassbarkeit der Daten erleichtern. Dafür wurde die Form der bestehenden Angaben verändert. Beispielsweise wurden schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht, Zahlenangaben rechtsbündig in die dafür vorgesehenen Kästchen eingetragen, verbale Angaben von Noten in Ziffern übersetzt (z. B. „gut“ = 2,0).

Zum anderen zielte die manuelle Prüfung darauf ab, schon vor der softwaregestützten Korrektur (siehe unten) erste Fehler oder Inkonsistenzen in den Angaben der Befragten zu bereinigen. Beispielsweise musste die Angabe „keine“ gestrichen werden, wenn ein oder mehr Aufsätze in einer Zeile bei „Anzahl insgesamt“ genannt wurden. Ferner wurde geprüft, ob die Angaben über Berufstätigkeiten im Tätigkeitstableau mit den entsprechenden Angaben zur Erwerbstätigkeit übereinstimmten (Welle 1, Fragen 2.1 und 8.1-8.13). Festgestellte Inkonsistenzen wurden – falls möglich – durch den Abgleich mit anderen Nennungen im Fragebogen aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.7) vergeben.

[Softwaregestützte Korrektur] Im Anschluss an die Datenübertragung in der ersten Welle erfolgte eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software und ab der zweiten Welle mit Hilfe des Statistikprogrammes Stata. Dabei sollten zum einen Fehler bei der

¹¹ Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

vorherigen manuellen Vorkorrektur und Datenübertragung, zum anderen weitere inkonsistente Angaben der Befragten, die bei der Vorkorrektur nicht geprüft werden konnten, identifiziert werden. Anhand formaler Regeln wurden gültige Wertebereiche und Antwortkombinationen definiert und geprüft. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variable in dem für diese Variable definierten Wertebereich lag.
- *Prüfung der Einhaltung der Filterführung:* Auf Grundlage der definierten Filterführung des Fragebogens wurde auf der einen Seite geprüft, ob für die jeweilige befragte Person Angaben zu erwarten gewesen wären, die aber nicht vorhanden waren (Vollständigkeitsprüfung) und auf der anderen Seite, ob für die jeweilige Person Angaben vorhanden waren, die nicht hätten erfolgen dürfen (Filterverstöße).
- *Prüfung von Merkmalskombinationen:* Es wurde die Konsistenz der Angaben innerhalb eines Fragebogens sowie zwischen den Befragungswellen geprüft. Detaillierte Informationen zur Datenedition des Primärforschungsteams finden sich in den Dateneditionskonzepten und im Datenversionierungskonzept.

Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens bzw. des Befragungsdatensatzes geprüft, ob der entsprechende Wert falsch (bzw. nicht) übertragen worden war. Ansonsten wurde versucht, den korrekten Wert anhand anderer Angaben im Fragebogen zu erschließen. Im Zweifelsfall wurde ein spezifischer Missingcode vergeben (vgl. Kapitel 6.7). Fehlerkorrekturen wurden dokumentiert und von mindestens einer weiteren Person geprüft.

[Löschung von Fällen] Im Panelverlauf wurden einige Fälle aus dem Datensatz entfernt. Zwei Fälle wurden gelöscht, weil eine doppelte Teilnahme festgestellt wurde. Löschungen wurden ebenfalls vorgenommen, wenn weniger als eine Frage beantwortet wurde (fünf Fälle) oder wenn zu viele Inkonsistenzen vorhanden waren (drei Fälle). Dreimal musste der komplette Fall aus dem Datensatz genommen werden, da ein Widerruf vorlag.

6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz des Promoviertenpanels 2014 auch generierte Variablen. Dabei handelt es sich zum einen um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 6.2). Zum anderen wurden Variablen für den SUF aus Datenschutzgründen verändert (vgl. Kapitel 8). Außerdem wurden im Forschungsfeld häufiger benötigte Variablen aus den Werten einer oder mehrerer Quellvariablen generiert (z. B. Ableitung der Kinderzahl aus den Angaben im Tableau zu Angaben über Kinder).

6.5 Erstellung der Datensätze

[Zusammenführung der Wellen] Die Daten aller Befragungswelle wurden zwar separat erhoben, jedoch in einem Datensatz zusammengeführt. Die Zuordnung der Fälle erfolgte über die im Rahmen der ersten Feldphase vergebenen Identifikationsnummern der Befragten (vgl. Kapitel 4).

[Erstellung von Personen- und Episodendatensatz] Die so zusammengeführten Daten wurden vom FDZ für den SUF in zwei getrennten Datensätzen abgelegt. Der Personendatensatz enthält den Großteil der Befragungsdaten sowie die zusätzlich generierten Variablen. Für jede befragte Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen.

Der Episodendatensatz im long-Format basiert weitestgehend auf den Informationen der Kalendarien, welche ab der zweiten Welle monatsgenaue Angaben zu den Tätigkeiten der Befragten liefern (sh. z. B. Frage 4.8, Welle 2). Dabei wurde in Welle 2 nach den Tätigkeiten seit Abschluss der Promotion bis zum Fragezeitpunkt gefragt, ab Welle 3 wurde der letzte Befragungszeitpunkt der teilnehmenden Person verwendet. Zusätzlich wurde in allen Wellen ein Berufstableau verwendet (sh. z. B. Frage 2.3, Welle 1), in dem die Befragten detaillierte Angaben zu ihren beruflichen Tätigkeiten machen konnten. Da beide Instrumente Informationen zu beruflichen Aktivitäten enthalten, wurden diese im Primärforschungsprojekt plausibilisiert.

Eine Ausnahme bildet hierbei die Kategorie „Selbstständigkeit/Lehrauftrag/Werkvertrag/Honorararbeit“ in der Variable ‚status‘ im Episodendatensatz. Diese basiert ausschließlich aus Informationen aus dem Berufstableau. Falls Informationen aus Welle 1 vorlagen, die befragte Person aber nicht an der zweiten oder dritten Welle teilgenommen hatte, so wurden die Informationen hier unter der Kategorie „Selbstständigkeit/Lehrauftrag/Werkvertrag/Honorararbeit“ zusammengefasst, sofern im Kalendarium keine gegenteiligen Informationen angegeben wurden. Da das Tableau in den ersten Wellen nicht klar zwischen Selbstständigkeit und Werkvertrag differenzieren kann, wurde diese hier kombiniert.

[Dateiformat] Alle Datensätze des FDZ-DZHW werden sowohl im Stata- als auch im SPSS-Format bereitgestellt.

6.6 Vergabe von Variablennamen, Variablen- und Wertelabels

[Variablen- und Wertelabelvergabe] Für Variablen- und Wertelabels wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabels in der Regel auf dem entsprechenden Fragetext. Grundlage für die Wertelabels sind je nach Fragetyp die Texte der Antwortoptionen bzw. eine Kombination der Texte von Frage und Antwortoption. Bei generierten Variablen, denen bestimmte Klassifikationen zugrunde liegen, wurden für die Wertelabels die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen. Mit dem Suffix „_g#“ werden darüber hinaus generierte Variablen (vgl. Kapitel 6.4) gekennzeichnet

[Variablenbenennung] Vom Primärforschungsteam wurden die Variablennamen nach einem festen Schema vergeben. Der erste Buchstabe des Variablennamens kennzeichnet den Themenbereich, dem eine Variable zugeordnet werden kann (Beispiel: „c“ bedeutet Themenbereich „Beschäftigungsverlauf, Tätigkeitsmerkmale“). Die zweite Stelle kennzeichnet die Welle, in der die Variable erhoben wurde.¹² Sofern eine Variable einem bestimmten Variablenblock zugeordnet werden kann, wurde ein Variablenstamm vergeben.¹³ Ein sprechender Namenszusatz gibt schließlich den konkreten Inhalt der jeweiligen Variable wieder.¹⁴ Textvariablen zu offenen Fragen wurden mit der Endung „_txt“ gekennzeichnet. Eine Auflistung der Themenbereiche und Variablenstämme, die für die Erstellung der Variablennamen verwendet wurden, befindet sich in Tabelle 4.

Tabelle 4: Systematik der Themenbereiche und Variablenstämme

Themenbereiche	Variablenstämme
A	Vorhandene Kompetenzen
B	Studienverlauf
	mop = Abschlussmonat jp = Abschlussjahr

¹² Beispiel: c3 bedeutet, dass die Variable aus der dritten Welle stammt

¹³ Beispiel: c3j bedeutet, dass die Variable zum Block „Aktuelle/letzte Stelle“ gehört

¹⁴ Beispiel: c3jbranche kennzeichnet die Variable „Welchem Wirtschaftsbereich gehören/gehörten der Betrieb bzw. die Einrichtung, in dem/der Sie arbeiten/arbeiteten schwerpunktmäßig an?“

		fachsem = Fachsemester examnote = Examensnote punkte = Punktzahl abs = Abschlussart studber = Studienbereich hs = Hochschule
C	Beschäftigungsverlauf; Tätigkeitsmerkmale	job = Angaben im Tätigkeitstableau zu = Empfehlungen für berufliche Zukunft ak = Wissenschaftliche Aktivitäten ko = Kontakte neustell = Stelle nach Abschluss der Prom. e = Erste Stelle j = Aktuelle/letzte Stelle zuf = Bereichsspezifische Zufriedenheit stud = Tätigkeiten zu Studienzeiten weib = berufsständische Weiterbildung ww = wissenschaftliche Weiterqualifizierungen beaus = antizipierte Beschäftigungsaussichten
D	Promotionsverlauf; Rahmenbedingungen der Promotion	pv = Promotionsvereinbarungen pub = wissenschaftliche Publikationen kon = Konferenzen/Tagungen mob = Mobilitätserfahrungen pat = Patente nw = Netzwerke bpraxis = Berufspraxis außerhalb der Wiss. fin = Finanzierung
K	Soziales Umfeld (Partnerschaft, Kinder, Demografie, Schulabschluss und soziale Herkunft)	geb = Geburt (Datum/Ort) part = Partnerschaft kind = Kinder studb = Studienberechtigung nat = Staatsangehörigkeit
P	Persönlichkeitseigenschaften, Lebensziele und Motive	m = Promotionsmotive asku = Allgemeine Selbstwirksamkeit (Kurzskala) kont = Kontrollüberzeugungen lz = Wichtigkeit Lebensziele bfi = Big-Five-Inventory mob = Mobilitätsbereitschaft
S	SSCO-Dimensionen	strs = Structure_Sicherheit und Stabilität strr = Structure_Regelungen und Kontrolle strg = Structure_Grundlgn./Ausstattung /Beding./Gelegenheitsstruk. chak = Challenge_Kooperatives Forschen chal = Challenge_Leistungsdruck allgemein chav = Challenge_Veröffentlichungsdruck orid = Orientation_Interdisziplinarität orii = Orientation_Internationalität supf = Support_Fachliche Unterstützung supe = Support_Emotionale Unterstützung supk = Support_Karriereplanung supw = Support_Wissenschaftl. Aktivitäten sups = Support_Schlüsselkompetenzen supa = Support_Ausbau wissenschaftlicher Netzwerke orif = Orientation_Forschungsorient. orip = Orientation_Praxisorient.
W	Wissenschaftliche Aktivitäten nach der Promotion	pub = wissenschaftliche Publikationen kon = Konferenzen pat = Patente wisskar = wissenschaftliche Karriere prof = Professur gut = Gutachtertätigkeiten foan = Anträge zur Forschungsförderung

		mitgl = Mitgliedschaften in Berufsverbänden/Fachgesellschaften asv = akademische Selbstverwaltung
G	Nicht-Erwerbstätigkeit	
M	Mobilität	plan = Mobilitätspläne
X	Informationen zum Fragebogen	

6.7 Codierung fehlender Werte

Zur Codierung fehlender Werte wird im FDZ-DZHW eine übergreifende Systematik genutzt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben wurden dabei durch dreistellige negative Werte codiert. Tabelle 5 stellt die verwendete Missingsystematik dar.

Die fehlenden Angaben lassen sich drei verschiedenen Gruppen zuordnen: In den ersten beiden Gruppen wird zwischen fehlenden Werten aufgrund von Nicht-Beantwortung von Fragen seitens der Befragten (Nonresponse) und fehlenden Werten aufgrund der Filterführung bzw. Irrelevanz der Frage für die Befragten unterschieden (nicht zutreffend). Die dritte Gruppe beinhaltet Missingcodierungen, die durch das Primärforschungsprojekt oder das FDZ im Zuge der Datenaufbereitung vergeben wurden (editierter fehlender Wert). Zu dieser Gruppe gehört auch die Codierung, die aufgrund von Anonymisierungsmaßnahmen (vgl. Kapitel 8) für bestimmte Variablen vergeben wurde.¹⁵

Tabelle 5: Systematik des FDZ-DZHW für fehlende Werte

Wertebereich	Code	Wertelabel
-999 bis -990: Nonresponse	-999 -998 -997 -996 -995 -994	weiß nicht keine Angabe keine Angabe (Antwortkategorie) Interviewabbruch keine Teilnahme (Panel) verweigert
-989 bis -970: Nicht zutreffend	-989 -988 -987 -986 -985	filterbedingt fehlend trifft nicht zu designbedingt fehlend (Fragebo- gensplit) designbedingt fehlend (Welle) ^a designbedingt fehlend (Kohorte) ^b
-969 bis -950: Editierter fehlender Wert	-969 -968 -967 -966 -965	unbekannter fehlender Wert ^c unplausibler Wert ^d anonymisiert nicht bestimmbar ^e ungültige Mehrfachnennung
-949 bis -930: Item-spezifische fehlende Werte	-949	nicht gegeben
-929 bis -920: Andere fehlende Werte	-929	Datenverlust

a Dieser Wert ist nur für Datensätze im Long-Format vergeben.

b Dieser Wert wird nur in gepoolten Datensätzen vergeben.

c Dieser Wert wird vergeben, wenn keinerlei Ursache rekonstruiert werden kann.

¹⁵ Eine vierte Gruppe umfasst spezielle Missingcodierungen, die im Rahmen der Datenaufbereitung eines konkreten Datensatzes nur für einzelne Items vergeben wurden.

- d Angaben, die aufgrund unterschiedlicher Faktoren in der Codierphase als nicht plausibel eingestuft werden, erhalten diesen Wert. Eine exakte Rekonstruktion ist ggf. nicht mehr möglich.
- e Diese Kategorie wird vergeben, wenn eine eindeutige Codierung nicht möglich ist, z. B. bei einer offenen Angabe, die nicht vercodet werden konnte, da sie nicht lesbar ist.

7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen im Vergleich zur definierten Grundgesamtheit. Es folgt zunächst eine allgemeine Einführung in die Vorgehensweise und eine Darstellung der erstellten Gewichte. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben.

In den Wellen 8-10 wurde aus Ressourcengründen auf eine Erstellung der Gewichtungsvariablen verzichtet.

7.1 Vorgehen und Anwendungshinweise

[Ursachen für die Verzerrungen der Stichproben] Maßgeblich für die Verzerrungen von Stichproben sind in der Regel zwei Prozesse:

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen.
- **Verzerrung durch Nonresponse:** Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe (vgl. Kapitel 5). Wenn diese Ausfallsprozesse unsystematisch sind (*Missing Completely at Random*), können sie ignoriert werden. Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (*Missing at Random*, *Missing Not at Random*), der einer Modellierung bedarf.¹⁶

[Konzeptuelles Vorgehen] Da es sich bei dem DZHW-Promoviertenpanel um eine Vollerhebung handelt, wurden keine designbedingten Disproportionalitäten erzeugt, die ausgeglichen werden müssten. Designgewichte mussten folglich nicht erstellt werden, da alle Individuen die gleiche Auswahlwahrscheinlichkeit aufweisen. Für diesen Datensatz sollte jedoch eine Adjustierung mit Hilfe von Ausfallgewichten im Querschnitt erfolgen, die auf der Grundlage von Informationen über Teilnehmer*innen und Nichtteilnehmer*innen erzeugt werden.

In Zusammenarbeit mit Gewichtungsexpert*innen der Gesis und des DZHW wurde das Gewichtungsverfahren im Vergleich zum SUF der ersten beiden Erhebungswellen (DOI: 10.21249/DZHW:phd2014:2.0.0) optimiert. Weil das Studiendesign des Promoviertenpanels Ausfälle von Individuen in einzelnen Erhebungswellen zulässt, werden im vorliegenden SUF ausschließlich Querschnittsgewichte bereitgestellt. Dafür werden im ersten Schritt die wellenspezifischen Ausfallgewichte anhand eines logistischen Regressionsmodells mit Prädiktoren aus der Erstbefragung erstellt (propensity scores). Diese Ausfallgewichte werden im zweiten Schritt mithilfe eines Raking-Algorithmus auf die Merkmalsverteilung der Grundgesamtheit kalibriert. Dadurch ergeben sich im Vergleich insgesamt kleinere Gewichte mit geringeren Varianzen. Ausreißer i. S. v. sehr kleinen oder sehr großen Gewichten werden schließlich getrimmt.

Darüber hinaus wurde ein Fehler bei der Zuordnung von Promovierten von Berliner Hochschulen bei der Kalibrierung auf die Merkmalsverteilung in der Grundgesamtheit berichtigt. In Tabelle 6 sind die bereitgestellten Gewichte dargestellt.

¹⁶ Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin 1976.

Tabelle 6: Bereitgestellte Gewichte zum DZHW-Promoviertenpanel 2014

Variablenname	Beschreibung
x1gewi	Kalibriertes Gewicht für Querschnittsanalysen der 1. Erhebungswelle (getrimmt)
x2geaus	Ausfallgewicht für Querschnittsanalysen der 2. Erhebungswelle (getrimmt)
x2gewi	kalibriertes Ausfallgewicht (x2geaus) für Querschnittsanalysen der 2. Erhebungswelle (getrimmt)
x3geaus	Ausfallgewicht für Querschnittsanalysen der 3. Erhebungswelle (getrimmt)
x3gewi	kalibriertes Ausfallgewicht (x3geaus) für Querschnittsanalysen der 3. Erhebungswelle (getrimmt)
x4geaus	Ausfallgewicht für Querschnittsanalysen der 4. Erhebungswelle (getrimmt)
x4gewi	kalibriertes Ausfallgewicht (x4geaus) für Querschnittsanalysen der 4. Erhebungswelle (getrimmt)
x5geaus	Ausfallgewicht für Querschnittsanalysen der 5. Erhebungswelle (getrimmt)
x5gewi	kalibriertes Ausfallgewicht (x5geaus) für Querschnittsanalysen der 5. Erhebungswelle (getrimmt)
x6geaus	Ausfallgewicht für Querschnittsanalysen der 6. Erhebungswelle (getrimmt)
x6gewi	kalibriertes Ausfallgewicht (x6geaus) für Querschnittsanalysen der 6. Erhebungswelle (getrimmt)
x7geaus	Ausfallgewicht für Querschnittsanalysen der 7. Erhebungswelle (getrimmt)
x7gewi	kalibriertes Ausfallgewicht (x7geaus) für Querschnittsanalysen der 7. Erhebungswelle (getrimmt)

[Hinweise zur Anwendung der Gewichte] Bei den erstellten Gewichten handelt es sich um probability weights, die in Stata mit Hilfe .ado-spezifischer Optionen berücksichtigt werden können.¹⁷ Das Gewicht x1gewi dient der Gewichtung der ersten Erhebungswelle, die Gewichte x2geaus und x2gewi dienen der Gewichtung bei Analysen der Zweiten-Welle-Daten usw. Die bereitgestellten Gewichte sind nur für Querschnittsanalysen geeignet. Da Panelaus- und Wiedereintritte zwischen Wellen möglich sind, wurde auf die Bereitstellung von Längsschnittgewichten im Rahmen des SUF verzichtet. Diese können nutzerseitig jedoch selbst erstellt werden. Grundlegend ist zu beachten, dass Gewichte nur dann sinnvolle Korrekturgrößen darstellen, wenn das verwendete Analysemodell die zur Gewichtung herangezogenen Variablen enthält oder mit diesen in einem Zusammenhang steht. Aus diesem Grund müssen Gewichte immer mit Fokus auf die jeweilige Fragestellung verwendet werden. Im Folgenden wird die Vorgehensweise bei der Erstellung der Gewichte für die Wellen 1 bis 7 näher dargestellt.

7.2 Modellierung der Ausfallgewichte

[Propensity Score Matching] Für die Wellen 2 bis 7 wurden die Ausfallgewichte mithilfe von Propensity Scores berechnet.¹⁸ Hierzu wurde anhand eines Logit-Schätzmodells die individuelle Teilnahme-wahrscheinlichkeit einer Person für die entsprechende Welle vorhersagt (RP). In einem explorativen Vorgehen wurden individuelle Merkmale identifiziert, die einen Beitrag zur Erklärung der Teilnahme-wahrscheinlichkeit leisten.¹⁹ Um für jede Person die Teilnahmewahrscheinlichkeit schätzen zu können, wurden fehlende Werte bei allen Variablen als zusätzliche Kategorie in das Modell integriert. Maßgebliches Kriterium für die Modellierung der Ausfallprozesse war, Modelle mit größtmöglicher Erklärungskraft bei gleichzeitiger Sparsamkeit zu finden, die die Varianzerhöhung so gering wie möglich halten. Herangezogen wurden dabei ausschließlich Prädiktoren aus der Erstbefragung.²⁰ Aus dem Modell

¹⁷ Siehe hierzu die Stata-Hilfe (Befehl: help weights).

¹⁸ Das Verfahren entspricht seiner Logik nach dem Propensity Score Matching, das auf Rosenbaum und Rubin 1983 zurückgeht (siehe auch Blumenstiel und Gummer 2015).

¹⁹ Siehe hierzu im Anhang das Schätzmodell.

²⁰ Versuche, den Panelcharakter der vorliegenden Datenbasis zu berücksichtigen, z. B. indem auch die Teilnahme (ja/nein), der Teilnahmezeitpunkt oder Prädiktoren aus weiteren vorherigen Wellen in den Schätzmodellen berücksichtigt werden, erhöhten zwar die Erklärungskraft der Ausfallmodelle, mussten aufgrund zu großer Gewichte und Varianzen jedoch verworfen werden.

konnte dann in einem zweiten Schritt die bedingte Teilnahmewahrscheinlichkeit abgeleitet werden, deren Kehrwert das Ausfallgewicht ($x2_{geaus} \cdot x7_{geaus}$) für die jeweilige Welle darstellt: $w = 1/RP$.

[Schätzmodelle] Die Schätzmodelle, die zur Erstellung der Ausfallgewichte genutzt werden, beinhalten als Prädiktoren die Variablen *Geschlecht*, *Alter*, *Geburtsort im Ausland*, *Promotionsform*, *Promotionsfach*, *Promotionsnote*, *Faulheit (Big-5)* und *das Brutto-Einkommen*. Die Logit-Modelle der Wellen zwei bis sieben können dem Anhang in Tabelle 9, Tabelle 10, Tabelle 11 und Tabelle 12 entnommen werden.

7.3 Kalibrierung

[IPF-Raking-Verfahren] Die Ausfallgewichte der Wellen 2 bis 7 ($x2_{geaus} \cdot x7_{geaus}$) wurden mithilfe des IPF-Raking-Verfahrens (Iterative Proportional Fitting) an die Randverteilung bekannter Merkmale in der Grundgesamtheit angepasst.²¹ Da für das Raking-Verfahren Informationen zur Randverteilung der Merkmale ausreichen und keine Kreuztabellen notwendig sind, müssen kaum Zellen zusammengefasst und weniger Einzelfallentscheidungen getroffen werden. Die Anpassung der Gewichte folgt dabei einem iterativen Prozess, bis keine Verbesserung mehr möglich ist. Da für die Erstbefragung keine Ausfallgewichte erstellt wurden, erfolgt für diese ausschließlich die Kalibrierung.²²

[Merkmalsverteilung der Grundgesamtheit] Der Abgleich der Stichproben mit der Grundgesamtheit erfolgte anhand der Merkmale Studienbereich, Geschlecht und Region der Promotionshochschule (Ost/West). Die Informationen zur Grundgesamtheit wurden den Daten des Statistischen Bundesamtes entnommen (Datenportal ICE-Land; Bestand: 50001). Personengruppen, deren Anteil in der jeweiligen Welle nicht dem Anteil in der Grundgesamt entspricht, werden unter Verwendung des kalibrierten Ausfallgewichts ($x1_{gewi} \cdot x7_{gewi}$) so entsprechend hoch- bzw. runtergewichtet.

Im Prozedere der Erstellung der Gewichtungsvariablen für die sechste Welle gab es eine Besonderheit, denn im Datensatz gibt es zwei Personen, deren Promotionsfach keinem der Studienbereiche zugeordnet werden kann. Diese Kategorie ist laut amtlicher Statistik auch in der Grundgesamtheit vorhanden. In Welle 6 (und nur in Welle 6) hat jedoch keine dieser Personen teilgenommen, sodass diese Kategorie in der Tabelle der Grundgesamtheit theoretisch mit einer ähnlichen oder höher geordneten Kategorie zusammengefasst werden müsste. Diese Zuordnung war jedoch nicht möglich, da diese besondere Kategorie keiner anderen Kategorie zuzuordnen ist. Ein Ausschluss dieser Fälle aus der Grundgesamtheit war ebenfalls nicht möglich. Also entschied sich die Primärforschungsgruppe zu folgendem Vorgehen: Nach dem Zufallsprinzip wurde einer der beiden Fälle im Datensatz als teilnehmend umcodiert und in die Gewichtungsprozedur einbezogen. Anschließend wurde er wieder als nicht teilnehmend (zurück) codiert und die Gewichtungsvariable auf „Missing“ gesetzt. Dieses Verfahren hat zur Folge, dass der Durchschnittswert der Gewichtungsvariable $x6_{gewi}$ nicht exakt bei 1 liegt, sondern minimal nach unten abweicht. Es ist anzunehmen, dass die errechneten Gewichte der anderen Fälle ebenfalls minimal vom tatsächlichen Wert abweichen.

[Fehlende Werte] Bei der Verwendung der kalibrierten Ausfallgewichte gilt es zu beachten, dass Personen, die keine Angaben bei den für das Raking relevanten Variablen (Geschlecht, Studienbereich, Bundesland der Promotionshochschule) gemacht haben, keine Kalibrierung durchgeführt werden konnte. Dementsprechend erhalten diese Personen keine kalibrierten Ausfallgewichte. Die Fallzahlen (n) können bei Analysen unter Verwendung kalibrierter Gewichte demnach geringer ausfallen.

²¹ Das IPF-Raking Verfahren geht auf Deming und Stephan (1940) zurück (sh. auch Lomax und Norman (2016)) und wurde von Kolenikov (2014) im stata-.ado ipfraking implementiert.

²² Technisch erfolgte dies, indem für alle Teilnehmenden der Erstbefragung im ersten Schritt ein Ausfallgewicht mit dem Wert 1 erstellt wurde, das im zweiten Schritt dann anhand des Raking-Verfahrens kalibriert wurde.

7.4 Trimmung

Um Extremwerte zu beseitigen, wurden alle Gewichte einer Trimmung nach Potter (1990) (vgl. auch Valliant et al. 2013, S. 388f.) unterzogen. Dem Verfahren liegt die Annahme zugrunde, dass die Gewichte einer Wahrscheinlichkeitsverteilung (Betaverteilung) folgen. All jene Gewichte, die über dem 99-Prozent-Quantil liegen, werden auf diese Grenze trunziert. Der Überschuss jenseits der Trunkierung wird im Folgenden unter den verbleibenden Gewichten verteilt.

7.5 Verteilung der Gewichte

Die Verteilungsmaße der bereitgestellten Gewichte für die Wellen eins bis zehn lässt sich Tabelle 7 entnehmen:

Tabelle 7: Verteilung der Gewichte für Welle 1 bis 7

Gewicht	n	Min	Max	SD	Varianz
x1gewi	5.318	0,226	1,832	0,264	0,069
x2geaus	3.183	1,163	2,326	0,230	0,053
x2gewi	3.146	0,178	2,162	0,349	0,122
x3geaus	2.924	1,264	2,603	0,272	0,074
x3gewi	2.890	0,166	2,162	0,349	0,122
x4geaus	2.981	1,241	2,559	0,267	0,072
x4gewi	2.949	0,209	2,147	0,342	0,117
x5geaus	3.037	1,246	2,432	0,239	0,057
x5gewi	3.003	0,214	2,055	0,323	0,104
x6geaus	2.535	1,463	3,025	0,322	0,103
x6gewi	2.509	0,253	2,112	0,337	0,113
x7geaus	2.574	1,411	3,045	0,335	0,113
x7gewi	2.545	0,196	2,146	0,341	0,117

8 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten²³, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz (BDSG) in seiner Neufassung vom 30. Juni 2017.²⁴ Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden“ können (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO).

[Sicherstellung des Datenschutzes im FDZ-DZHW] Im FDZ-DZHW wird der Datenschutz der Befragten über rechtlich-organisatorische Maßnahmen, technische Maßnahmen und statistische Maßnahmen sichergestellt²⁵, die in unterschiedlicher Weise kombiniert werden können. Aus den Kombinationen ergeben sich vier potenzielle Zugangswege, die in Tabelle 8 im Überblick dargestellt sind.²⁶

Tabelle 8: Zugangswege im FDZ-DZHW

	Campus Use File (CUF): Download	Scientific Use File (SUF): Download	Scientific Use File (SUF): Remote-Desktop	Scientific Use File (SUF): On-Site
rechtlich-organisatorische Maßnahmen	sehr gering	gering	moderat	hoch
technische Maßnahmen	sehr gering	gering	moderat	hoch
statistische Maßnahmen	sehr hoch	hoch	moderat	gering

Je stärker der Datenzugang rechtlich-organisatorisch und technisch kontrolliert und beschränkt wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten. In der Folge müssen die Daten umso weniger mittels statistischer Maßnahmen anonymisiert, also um Informationen reduziert, werden – das heißt, desto größer bleibt ihr Analysepotenzial.

²³ „Personenbezogene Daten“ (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden ‚betroffene Person‘) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind“ (Art. 4 Abs. 1 DSGVO).

²⁴ Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni 2017 (Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (§ 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn.

²⁵ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (IfBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, 9f.) orientieren.

²⁶ Für genauere Informationen vgl. <https://www.fdz.dzhw.eu/de/datennutzung>

Für die Daten des Promoviertenpanel 2014 gibt es alle Zugangswege. Im Folgenden werden die konkret durchgeführten statistischen Anonymisierungsmaßnahmen genauer erläutert.

[Statistische Anonymisierungsmaßnahmen] Im Rahmen der statistischen Anonymisierungsmaßnahmen wurden zunächst alle Informationen darauf geprüft, ob sich über sie Personen direkt identifizieren lassen. Diese direkten Identifikatoren, wie Namen, Adressen und E-Mail-Adressen, wurden bereits während der Feldphase in einem separaten Datensatz erfasst und sind somit per se nicht in den Daten enthalten. Die Original-Identifikationsnummer wurde entfernt und durch eine neue zufällig vergebene Identifikationsnummer ersetzt.

Darüber hinaus wurde der Empfehlung von Ebel und Meyermann gefolgt, offene Angaben zu löschen „selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten“ (Ebel & Meyermann, 2015, S. 5).

Anschließend wurden die Quasi-Identifikatoren bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren. Um eine eindeutige Zuordnung der Daten zu unterbinden, werden diese Schlüsselmerkmale – je nach Zugangsweg – nur in aggregierter Form freigegeben oder komplett anonymisiert.

Zuletzt wurde geprüft, ob in den Daten sensible Informationen, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Diese eignen sich zwar nicht notwendig zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg, 2016, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Die sensiblen Informationen werden – je nach Zugangsweg – nur in aggregierter Form freigegeben oder komplett anonymisiert.

Im Datensatzreport, welcher im [Metadaten](#)system des FDZ-DZHW zu finden ist, sind weitere Informationen darüber, welche Variablen in den bereitgestellten Daten enthalten sind und über welche(n) Zugangsweg(e) sie jeweils nutzbar sind, aufgeführt. So kann das Bundesland des Arbeitsorts aggregiert zu Bundeslandgruppen (c*taet1bula_g1) über alle Zugangswege genutzt werden. Unaggregiert kann das Bundesland nur über die Zugangswege Remote-Desktop und On-Site genutzt werden (c*taet1bula).

9 Anhang

9.1 Logit-Regressionen zur Erstellung der Panelausfallgewicht

Tabelle 9: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 3

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,10	0,07	1,49	0,14
keine Angabe	1,13	0,51	0,28	0,78
Alter				
23-30	1,04	0,08	0,57	0,57
31-34	Ref.			
35-39	1,01	0,08	0,16	0,87
40-44	1,25	0,17	1,64	0,10
45-49	1,90	0,36	3,38	0,00
50-60	1,75	0,37	2,66	0,01
61-79	2,17	0,86	1,97	0,05
keine Angabe	0,12	0,07	-3,71	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,95	0,08	-0,56	0,58
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Pro- motionsprogramm	1,03	0,12	0,28	0,78
Stipendienpro- gramm	0,79	0,09	-2,14	0,03
frei promovierend	0,92	0,09	-0,91	0,36
keine Angabe	0,62	0,25	-1,17	0,24
Promotionsnote				
summa cum laude	1,32	0,13	2,89	0,00
magna cum laude	1,17	0,09	2,08	0,04
cum laude	Ref.			
satis bene	1,24	0,47	0,56	0,57
rite	0,96	0,20	-0,22	0,83
keine Angabe	0,52	0,17	-2,01	0,04

Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,68	0,07	-3,98	0,00
keine Angabe	0,90	0,21	-0,44	0,66
"Ich bin bequem, neige zu Faulheit"				
trifft überhaupt nicht zu	0,78	0,06	-3,21	0,00
2	0,90	0,07	-1,28	0,20
3	Ref.			
4	1,12	0,13	0,93	0,35
trifft voll und ganz zu	0,87	0,20	-0,61	0,54
keine Angabe	0,33	0,18	-2,01	0,05
Fächergruppe				
Geisteswissenschaften	1,19	0,14	1,43	0,15
Sport	0,68	0,23	-1,11	0,27
Rechts-, Wirtschafts- und Sozialwiss.	0,82	0,07	-2,23	0,03
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,73	0,07	-3,21	0,00
Agrar-, Forst-, Ernährungswiss, Veterinärmed.	0,91	0,13	-0,66	0,51
Ingenieurwissenschaften	0,82	0,08	-2,04	0,04
Kunst, Kunstwissenschaft	1,02	0,25	0,08	0,94
keine Angabe	1,06	0,43	0,15	0,88
aktuelles/letztes Bruttoeinkommen				
bis 3200€	1,01	0,08	0,11	0,91
3201€ bis 5000€	Ref.			
über 5000€	0,88	0,07	-1,63	0,10
kein Einkommen	0,83	0,10	-1,51	0,13
keine Angabe	0,56	0,06	-5,08	0,00
Konstante	1,45	0,17	3,16	0,00
N=5408; Pseudo R ² =0,025				

Tabelle 10: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 4

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,07	0,07	1,17	0,24
keine Angabe	0,83	0,37	-0,43	0,67
Alter				
23-30	1,05	0,08	0,70	0,48
31-34	Ref.			
35-39	0,89	0,07	-1,57	0,12
40-44	0,95	0,13	-0,36	0,72
45-49	1,81	0,34	3,13	0,00
50-60	1,52	0,32	2,03	0,04
61-79	1,25	0,47	0,60	0,55
keine Angabe	0,11	0,06	-3,87	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,93	0,08	-0,94	0,35
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Pro- motionsprogramm	0,95	0,11	-0,45	0,66
Stipendienpro- gramm	0,87	0,09	-1,26	0,21
frei promovierend	0,85	0,08	-1,78	0,08
keine Angabe	0,60	0,25	-1,25	0,21
Promotionsnote				
summa cum laude	1,27	0,12	2,44	0,02
magna cum laude	1,12	0,09	1,51	0,13
cum laude	Ref.			
satis bene	1,07	0,41	0,18	0,86
rite	0,75	0,16	-1,37	0,17
keine Angabe	0,59	0,19	-1,66	0,10
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,66	0,06	-4,43	0,00
keine Angabe	1,06	0,25	0,23	0,82

**"Ich bin bequem,
neige zu Faulheit"**

trifft überhaupt nicht zu	0,84	0,07	-2,26	0,02
2	0,95	0,07	-0,72	0,47
3	Ref.			
4	1,28	0,15	2,10	0,04
trifft voll und ganz zu	1,12	0,26	0,50	0,62
keine Angabe	0,68	0,35	-0,75	0,45

Fächergruppe

Geisteswissenschaf- ten	0,99	0,12	-0,09	0,93
Sport	0,71	0,24	-1,00	0,32
Rechts-, Wirtschafts- und Sozialwiss.	0,87	0,08	-1,59	0,11
Mathematik, Natur- wiss.	Ref.			
Humanmed./Ge- sundheitswiss.	0,69	0,07	-3,66	0,00
Agrar-, Forst-, Ernae- hungswiss, Veteri- närm.	0,90	0,13	-0,69	0,49
Ingenieurwissen- schaften	0,84	0,08	-1,76	0,08
Kunst, Kunstwissen- schaft	1,10	0,27	0,37	0,71
keine Angabe	1,07	0,44	0,17	0,87

**aktuelles/letztes
Bruttoeinkommen**

bis 3200€	1,02	0,08	0,31	0,76
3201€ bis 5000€	Ref.			
über 5000€	0,91	0,07	-1,18	0,24
kein Einkommen	0,95	0,12	-0,39	0,69
keine Angabe	0,55	0,06	-5,22	0,00
Konstante	1,61	0,19	4,04	0,00

N=5408; Pseudo
R²=0,026

Tabelle 11: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 5

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,05	0,07	0,85	0,40
keine Angabe	1,11	0,51	0,23	0,82
Alter				
23-30	1,05	0,08	0,67	0,50
31-34	Ref.			
35-39	0,99	0,08	-0,11	0,92
40-44	1,34	0,19	2,10	0,04
45-49	2,10	0,41	3,78	0,00
50-60	1,59	0,33	2,20	0,03
61-79	1,00	0,37	0,00	1,00
keine Angabe	0,08	0,05	-3,93	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,83	0,07	-2,32	0,02
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Pro- motionsprogramm	0,98	0,12	-0,20	0,84
Stipendienpro- gramm	0,85	0,09	-1,53	0,13
frei promovierend	0,89	0,08	-1,28	0,20
keine Angabe	0,55	0,23	-1,44	0,15
Promotionsnote				
summa cum laude	1,27	0,12	2,44	0,02
magna cum laude	1,18	0,09	2,15	0,03
cum laude	Ref.			
satis bene	0,89	0,34	-0,30	0,76
rite	1,17	0,24	0,75	0,45
keine Angabe	0,83	0,26	-0,61	0,54
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,72	0,07	-3,47	0,00
keine Angabe	0,93	0,22	-0,29	0,78
"Ich bin bequem, neige zu Faulheit"				

trifft überhaupt nicht zu	0,76	0,06	-3,47	0,00
2	0,83	0,06	-2,45	0,01
3	Ref.			
4	1,07	0,13	0,53	0,59
trifft voll und ganz zu	0,83	0,19	-0,79	0,43
keine Angabe	0,57	0,29	-1,12	0,26

Fächergruppe

Geisteswissenschaften	1,07	0,13	0,60	0,55
Sport	1,11	0,39	0,29	0,77
Rechts-, Wirtschafts- und Sozialwiss.	0,84	0,07	-2,03	0,04
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,73	0,07	-3,10	0,00
Agrar-, Forst-, Ernährungswiss, Veterinärw.	0,76	0,11	-1,94	0,05
Ingenieurwissenschaften	0,82	0,08	-2,05	0,04
Kunst, Kunstwissenschaft	0,89	0,22	-0,50	0,62
keine Angabe	1,22	0,51	0,48	0,63

aktuelles/letztes Bruttoeinkommen

bis 3200€	0,97	0,07	-0,37	0,71
3201€ bis 5000€	Ref.			
über 5000€	0,91	0,07	-1,26	0,21
kein Einkommen	0,89	0,11	-0,94	0,35
keine Angabe	0,50	0,06	-6,16	0,00
Konstante	1,77	0,21	4,86	0,00

N=5408; Pseudo
R²=0,025

Tabelle 12: Logit-Regression zur Erstellung des Ausfallgewichts für Welle 6

	Odds Ratio	Std. Err	z	P> z
Geschlecht				
Männlich	Ref.			
Weiblich	1,07	0,07	1,14	0,26
keine Angabe	0,78	0,35	-0,57	0,57
Alter				
23-30	1,07	0,08	0,96	0,34
31-34	Ref.			
35-39	0,91	0,07	-1,25	0,21
40-44	1,27	0,17	1,75	0,08
45-49	1,40	0,25	1,83	0,07
50-60	1,19	0,24	0,87	0,39
61-79 und keine Angabe	0,25	0,09	-3,88	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,89	0,07	-1,43	0,15
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Promotionsprogramm	1,09	0,13	0,76	0,45
Stipendienprogramm	0,91	0,10	-0,84	0,40
frei promovierend	0,95	0,09	-0,52	0,61
keine Angabe	0,86	0,35	-0,38	0,71
Promotionsnote				
summa cum laude	1,21	0,12	1,98	0,05
magna cum laude	1,15	0,09	1,79	0,07
cum laude	Ref.			
satis bene	0,85	0,33	-0,41	0,69
rite	1,24	0,26	1,04	0,30
keine Angabe	0,98	0,31	-0,07	0,94
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,69	0,07	-3,82	0,00
keine Angabe	0,75	0,18	-1,22	0,22
"Ich bin bequem, neige zu Faulheit"				

trifft überhaupt nicht zu	0,86	0,07	-2,00	0,05
2	1,00	0,08	0,02	0,99
3	Ref.			
4	1,32	0,15	2,42	0,02
trifft voll und ganz zu	1,01	0,23	0,04	0,97
keine Angabe	0,77	0,40	-0,50	0,62

Fächergruppe

Geisteswissenschaften	1,07	0,13	0,60	0,55
Sport	0,64	0,22	-1,29	0,20
Rechts-, Wirtschafts- und Sozialwiss.	0,81	0,07	-2,36	0,02
Mathematik, Naturwiss.	Ref.			
Humanmed./Gesundheitswiss.	0,76	0,08	-2,74	0,01
Agrar-, Forst-, Ernährungswiss., Veterinärmed.	0,79	0,11	-1,69	0,10
Ingenieurwissenschaften	0,88	0,09	-1,32	0,19
Kunst, Kunstwissenschaft	1,13	0,28	0,50	0,62
keine Angabe	0,89	0,35	-0,30	0,77

aktuelles/letztes Bruttoeinkommen

bis 3200€	1,01	0,08	0,17	0,87
3201€ bis 5000€	Ref.			
über 5000€	0,87	0,07	-1,76	0,08
kein Einkommen	0,84	0,10	-1,41	0,16
keine Angabe	0,60	0,07	-4,42	0,00
Konstante	1,04	0,12	0,32	0,75

N=5408; Pseudo R²=0,019

	Odds Ratio	Std. Err	z	P> z
--	------------	----------	---	------

Geschlecht

Männlich	Ref.			
Weiblich	1,06	0,06	0,94	0,35
keine Angabe	0,80	0,37	-0,48	0,63

Alter

23-30	1,18	0,09	2,21	0,03
31-34	Ref.			
35-39	0,96	0,07	-0,49	0,63
40-44	1,30	0,18	1,89	0,06

45-49	1,91	0,35	3,49	0,00
50-60	1,67	0,34	2,50	0,01
61-79	0,63	0,25	-1,17	0,24
keine Angabe	0,12	0,08	-3,33	0,00
Promotionsrahmen				
wiss. Mitarbeiter Haushaltstelle	0,95	0,08	-0,66	0,51
wiss. Mitarbeiter Drittmittelstelle	Ref.			
strukturiertes Pro- motionsprogramm	1,08	0,12	0,63	0,53
Stipendienpro- gramm	0,89	0,10	-1,10	0,27
frei promovierend	0,94	0,09	-0,63	0,53
keine Angabe	0,72	0,30	-0,79	0,43
Promotionsnote				
summa cum laude	1,26	0,12	2,36	0,02
magna cum laude	1,13	0,09	1,64	0,10
cum laude	Ref.			
satis bene	0,68	0,27	-0,96	0,34
rite	1,19	0,25	0,84	0,40
keine Angabe	1,17	0,36	0,50	0,62
Geburtsort				
Deutschland	Ref.			
In einem anderen Land	0,66	0,06	-4,25	0,00
keine Angabe	0,96	0,23	-0,17	0,86
"Ich bin bequem, neige zu Faulheit"				
trifft überhaupt nicht zu	0,83	0,06	-2,42	0,02
2	0,98	0,08	-0,26	0,79
3	Ref.			
4	1,35	0,16	2,61	0,01
trifft voll und ganz zu	0,93	0,21	-0,34	0,74
keine Angabe	0,39	0,23	-1,62	0,11
Fächergruppe				
Geisteswissenschaf- ten	1,05	0,12	0,38	0,70
Sport	0,68	0,24	-1,10	0,27
Rechts-, Wirtschafts- und Sozialwiss.	0,85	0,07	-1,85	0,07
Mathematik, Natur- wiss.	Ref.			

Humanmed./Gesundheitswiss.	0,72	0,07	-3,26	0,00
Agrar-, Forst-, Ernährungswiss, Veterinärn.	0,81	0,12	-1,46	0,14
Ingenieurwissenschaften	0,81	0,08	-2,15	0,03
Kunst, Kunstwissenschaft	1,01	0,24	0,03	0,98
keine Angabe	1,21	0,49	0,48	0,63
aktuelles/letztes Bruttoeinkommen				
bis 3200€	1,03	0,08	0,37	0,710
3201€ bis 5000€	Ref.			
über 5000€	0,93	0,07	-0,96	0,34
kein Einkommen	0,88	0,11	-1,09	0,28
keine Angabe	0,60	0,07	-4,38	0,00
Konstante	1,04	0,12	0,30	0,77
N=5408; Pseudo R2=0,022				

9.2 Dokumentation der Herkunft sekundär genutzter Quellen

Aufgrund des Umfangs erfolgt die Dokumentation der Herkunft sekundär genutzter Fragen in einer separaten Datei. Diese ist [hier](#) abrufbar.

10 Literaturverzeichnis

- Beierlein, C., Kovaleva, A., László, Z., Kempler, C.J. & Rammstedt, B. (2014a). Eine Single-Item-Skala zur Erfassung der Allgemeinen Lebenszufriedenheit: Die Kurzskaala Lebenszufriedenheit-1 (L-1). Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS-Working Papers, 2014|33).
- Beierlein, C., Kovaleva, A., Kemper, C.J. & Rammstedt, B. (2014b). Allgemeine Selbstwirksamkeit Kurzskaala (ASKU). Zusammenstellung sozialwissenschaftlicher Items und Skalen. Hg. v. GESIS – Leibniz-Institut für Sozialwissenschaften.
- Blumenstiel, J. E. & Gummer, T. (2015). Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In: Jürgen Schupp und Christof Wolf (Hg.): Non-response Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen. Wiesbaden: Springer Fachmedien Wiesbaden, S. 13–44.
- Brandt, G., de Vogel, S. & Jaksztat, S. (2016). Entwicklung und Testung eines Instruments zur Erfassung der Lernumwelt in der Promotionsphase. Ergebnisse der Entwicklungsstudie. Werkstattbericht. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. Hannover.
- Centers for Disease Control and Prevention (CDC) (2000). Measuring Healthy Days: Population Assessment of Health-Related Quality of Life. Atlanta, Georgia: CDC. Online verfügbar unter https://archive.cdc.gov/www_cdc_gov/hrqol/pdfs/mhd.pdf, zuletzt geprüft am 07.07.2025.
- de Vogel, S., Brandt, G. & Jaksztat, S. (2017). Ein Instrument zur Erfassung der Lernumwelt Promotionsphase. In: Zeitschrift für empirische Hochschulforschung (1), S. 24–44.
- Deming, W. E., Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. In: Ann. Math. Statist. 11 (4), S. 427–444. DOI: 10.1214/aoms/1177731829.
- Ebel, T. (2015). Empfehlungen zur Anonymisierung quantitativer Daten. GESIS - Leibniz-Institut für Sozialwissenschaften. Mannheim.
- Ebel, T., & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. Forschungsdaten Bildung informiert. Bd.3. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. <https://doi.org/10.25656/01:21970> (Zugegriffen am: 14.10.2024).
- Hochfellner, D., Müller, D., Schmucker, A., & Roß, E. (2012). Datenschutz am Forschungsdatenzentrum. (FDZ-Methodenreport 06/2012). Nürnberg: Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (IAB). https://doku.iab.de/fdz/reporte/2012/MR_06-12.pdf (Zugegriffen am: 15.07.2025).
- Jaksztat, S. (2017). Geschlecht und wissenschaftliche Produktivität. In: Zeitschrift für Soziologie 46 (5), S. 347–361.
- Jaksztat, S., Brandt, G., de Vogel, S. & Briedis, K. (2017). Gekommen, um zu bleiben? Die Promotion als Wegbereiter wissenschaftlicher Karrieren. In: WSI-Mitteilungen 70 (5), S. 321–329.
- Kobarg, S., Wollersheim, J., Welp, I. M., Spörrle, M. (2017). Individual Ambidexterity and Performance in the Public Sector: A Multilevel Analysis. In: International Public Management Journal 20 (2), S. 226–260. DOI: 10.1080/10967494.2015.1129379.

- Koberg, T. (2016). Disclosing the National Educational Panel Study. In: Hans-Peter Blossfeld, Jutta von Maurice, Michael Bayer und Jan Skopek (Hg.): *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study*. Wiesbaden: Springer VS, S. 691–708.
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). In: *The Stata Journal* 14 (1), S. 22–59.
- Kovaleva, A., Beierlein, C., Kemper, C. J. & Rammstedt, B. (2014). *Internale-Externale-Kontrollüberzeugung-4 (IE-4). Zusammenstellung sozialwissenschaftlicher Items und Skalen*. Hg. v. GESIS - Leibniz-Institut für Sozialwissenschaften.
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. In: *Transactions on Data Privacy* 1 (1), S. 2–16.
- Lomax, N. & Norman, P. (2016). Estimating Population Attribute Values in a Table: “Get Me Started in” Iterative Proportional Fitting. In: *The Professional Geographer* 68 (3), S. 451–461. DOI: 10.1080/00330124.2015.1099449.
- Mom, T. J. M., van den Bosch, F. A. J. & Volberda, H. W. (2009). Understanding Variation in Managers’ Ambidexterity: Investigating Direct and Interaction Effects of Formal Structural and Personal Coordination Mechanisms. In: *Organization Science* 20 (4), S. 812–828. DOI: 10.1287/orsc.1090.0427.
- Morgeson, F. P. & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. In: *The Journal of applied psychology* 91 (6), S. 1321–1339. DOI: 10.1037/0021-9010.91.6.1321.
- Nationales Bildungspanel (NEPS) (2013). Startkohorte 6: Erwachsene (SC6). Wellen 2 und 3. Erhebungsinstrumente (Feldversion). Bamberg: IflBi.
- Nationales Bildungspanel (NEPS) (2018). Erhebungsinstrumente (Feldversion). NEPS Startkohorte 5 — Studierende. Hochschulstudium und Übergang in den Beruf, Welle 10. Bamberg: IflBi.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In: *Proceedings of the Survey Research Methods Section*, S. 225–230.
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). *Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen*. Hg. v. GESIS – Leibniz-Institut für Sozialwissenschaften.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. In: *Biometrika* 70 (1), S. 41–55.
- Rubin, D. B. (1976). Inference and missing data. In: *Biometrika* 63 (2), S. 581–592.
- Scherer, S. & Brüderl, J. (2010). Sequenzdatenanalyse. In: Christof Wolf und Henning Best (Hg.): *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 1031–1051.
- Schilke, O. (2014). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. In: *Strat. Mgmt. J.* 35 (2), S. 179–203. DOI: 10.1002/smj.2099.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung*. 7. Aufl. München: Oldenbourg.

Statistisches Bundesamt (2015): Bildung und Kultur. Prüfungen an Hochschulen 2014. Hg. v. Statistisches Bundesamt. Statistisches Bundesamt. Wiesbaden.

Stegmann, S., van Dick, R., Ullrich, J., Charalambous, J., Menzel, B., Egold, N. & Wu, T. T.-C. (2010). Der Work Design Questionnaire. In: Zeitschrift für Arbeits- und Organisationspsychologie A&O 54 (1), S. 1–28. DOI: 10.1026/0932-4089/a000002.

Valliant, R., Dever, J. A. & Kreuter, F. (2013). Practical tools for designing and weighting survey samples. New York (NY): Springer New York.